

**UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF NEW JERSEY**

|   |   |
|---|---|
| PEARSON EDUCATION, INC.,<br><br><p style="text-align: center;">Plaintiff,</p> <p style="text-align: center;">v.</p> CHEGG, INC.,<br><br><p style="text-align: center;">Defendant.</p> | Civil Action No.: 21-16866<br><br><br><p style="text-align: center;"><b>OPINION AND ORDER OF THE<br/>SPECIAL DISCOVERY MASTER RE:<br/>PLAINTIFF’S REQUEST FOR<br/>PRODUCTION OF OUTSTANDING<br/>DOCUMENTS</b></p> |
|---|---|

**LINARES, J.**

This matter comes before the Special Master by way of Plaintiff Pearson Educations, Inc.’s letter application to compel Defendant Chegg, Inc. to produce additional data fields from its Textbook Solutions database. (ECF No. 106 at 3). Pearson simultaneously opposed Chegg’s application in the aforementioned joint letter. (*See generally* id.). The Special Master held Oral Argument on April 4, 2023 with respect to outstanding discovery disputes.<sup>1</sup> During that hearing, the Special Master requested supplemental briefing from the parties. Pearson made its supplemental submission on April 14, 2023 and Chegg made its supplemental responsive submission on April 21, 2023. For the reasons set forth below, the Special Master hereby **GRANTS** in part and **DENIES** in part Pearson’s application.

---

<sup>1</sup> The Transcript from the April 4, 2023 Hearing shall be cited as “Hrg. Tr.” followed by the pertinent page and line numbers.

## **I. BACKGROUND**

The Special Masters presumes the parties familiarity with the factual background and procedural posture of this matter. Accordingly, the Special Master will only recite the facts relevant to the disposition of the subject dispute.

This dispute involves Pearson's request for information contained in additional fields from Chegg's Textbook Solutions database ("TBS Database"). Pearson claims that in February 2022, it requested certain documents that go to the heart of Pearson's infringement claims concerning Chegg's creation and sale of unauthorized solutions to end-of-chapter questions in Pearson's copyrighted works. (*See* Pearson's April 14, 2023 Supplemental Submission at 1). According to Pearson, RFPs 1, 4, 5, 7, 9, 10, 11, 13, 14, 18, 19, 20, and 22 ("Subject RFPs") were all directed to narrow categories of documents. (*Id.*). Pearson advances that Chegg admits that its TBS Database tracks detailed information for each solution on Chegg Study, including for the 150 works at issue in this case. (*Id.* at 2). Pearson claims that Chegg confirmed that its TBS Database contains "all of the relevant and responsive information . . . associate[d] with the Pearson textbooks on Exhibit A to the Complaint that are or have been on Textbook Solutions" and that Chegg agreed to produce same. (*Id.*); (Hrg. Tr. 47:6-21).

During the April 4, 2023 hearing, Pearson explained that the demanded discovery is relevant because it represents "the actual questions and answers" in Chegg's TBS Database, and includes relevant information about who drafted the solutions, when they were drafted, and whether the solutions were "rejected as having a plagiarism problem when they were posted." (Hrg. Tr. 45:5-8).

Pearson further contends that on June 24, 2022, Chegg submitted a letter in which it claimed to have conducted a reasonable search of its database in an effort to identify data fields

that could correspond to Pearson's discovery requests. (Hrg. Tr. 47:22-48:6). Accordingly, Chegg provided Pearson with data from a subset of fields from the TBS Database (21 fields). (Hrg. Tr. 48:7-9). Upon reviewing this subset, Pearson requested that Chegg provide it with three sample exports that identified all available fields in the TBS Database (Id. at 48:19-21). Pearson then requested that Chegg provide the same information as to all available fields (minus a handful) as to the remaining 147 works in suit. (Hrg. Tr. 49:5-9; *see also* Pearson's April 14, 2023 Supplemental Submission at 2).

According to Pearson, these 250+ additional fields from the TBS Database exports will contain the following information which Pearson claims is critical to its case: (1) text of the Pearson questions answered on TBS; (2) results of Chegg's review and/or quality control process (if any); (3) Pearson's copyrighted content Chegg used or considered in writing TBS answers; (4) identity of persons involved in answering Pearson's questions on TBS; (5) extent of Chegg's distribution of the infringing answers on TBS; and (6) timeline of each TBS answer. (*See* Pearson's April 14, 2023 Supplemental Submission at 2-4).

In response, Chegg explains that the TBS Database was created by a company Chegg acquired in 2010 which was then further developed by Chegg, but as a result, there are fields in the database that have never been used by Chegg, and either have no data or data that predates Chegg and have nothing to do with the issues in the case. (*See* Chegg's April 21, 2013 Supplemental Submission at 1). Chegg counters that it conducted a reasonable search of the TBS fields and produced data for a subset of 21 fields in the TBS Database for all 150 titles and is further prepared to produce additional data from a limited number of additional fields. (Id. at 1; *see also* Hrg. Tr. 50:22-24). The 21 fields Chegg produced included the ISBN, title, author, problem number, problem text, solution text to make that comparison, relevant dates and time

stamps reflecting when the solution was created and when it was uploaded. (Hrg. Tr. 50:24-51:4). For the initial export of 21 fields for all 150 works, Chegg states that it took 150 hours of engineering time. (Chegg April 21, 2023 Supplemental Submission, at 2).

Chegg avers that notwithstanding its efforts, Pearson is never satisfied with the production, continually requesting additional information without necessarily understanding what it is seeking. (Hrg. Tr. 51:6-8). Pearson, Chegg claims, has “articulated no particular reason why [it needs] the vast majority of these other fields.” (Id. at 6-10). Rather, Chegg insists that Pearson’s request is simply an attempt to engage in a fishing expedition into its TBS Database, which it asserts is proprietary and contains “hundreds of tables,” each of which “could have dozens of fields.” (Hrg. Tr. 52:6-8, 57:24-25). According to Chegg, this translates into “many hundreds or thousands of fields for every single piece of information in the database.” (Hrg. Tr. 58:1-5).

Chegg nonetheless recognizes that certain other fields in the TBS Database may contain responsive information and is therefore amenable to producing “limited additional fields” of data “in a different file format that Chegg anticipates would make working with the files easier for Pearson.” (See Chegg’s April 21, 2023 Supplemental Submission at 1, 4). Specifically, Chegg proposes that it produce data from specific in-use fields relating to quality control and date information. (Id. at 4-6). Finally, Chegg agrees to produce data from a discrete set of additional identified fields, to the extent that they are actually in use in the TBS Database. (Id. at 6).

## **II. ANALYSIS**

The discovery dispute at issue has been ongoing since at least February 2022, during which time the parties have been actively negotiating the scope of discovery, and has been the subject of two hearings before the Court. While the Special Master understands Pearson’s position and agrees that it is entitled to certain additional information from the TBS Database, the Special

Master agrees with Chegg that to require production of data for an additional 250+ fields for the remaining 147 works at issue would be unduly burdensome and disproportionate to the needs of the case, as many of the fields do not contain information that is responsive to Pearson's requests. The Special Master will address each of the categories of information requested by Pearson below.

With respect to Pearson's request for Chegg to produce the text of Pearson's questions in a "readable" format, Chegg has agreed to produce the data in a different format that it anticipates will be easier for Pearson to understand.

With respect to Pearson's request for additional fields regarding quality control, Chegg has agreed to produce data from nine additional quality control fields that exist and are "in use." (Chegg April 21, 2023 Supplemental Submission at 4). The Special Master accepts Chegg's representation that certain of the fields Pearson has requested simply do not exist in the TBS Database. The Special Master also accepts Chegg's representation that many of the fields "have never been used by Chegg." Pearson is free to inquire into Chegg's statements during depositions, and, if that turns out not to be the case, Pearson may raise this issue with the Special Master. For purposes of this application, the Special Master understands that Chegg's use of the phrase "in use" does not mean *presently* in use, but rather, used at any point in time by Chegg. Stated differently, the limitation by Chegg to produce "in use" fields excludes from the production only fields that have *never* been used by Chegg. Accordingly, Chegg shall produce data from the nine additional fields set forth in its April 21, 2023 Supplemental Submission at page 4, to the extent those fields exist and have been used by Chegg.

With respect to Pearson's request for additional fields regarding content used by Chegg in preparing its solutions to Pearson's questions, the Special Master finds that this request seeks

information that is relevant and discoverable. Accordingly, Chegg shall produce data from those five additional fields identified by Pearson in its April 14, 2023 Supplemental Submission at page 3, to the extent those fields exist and have been used at any time by Chegg.

With respect to Pearson's request for identifying information of individuals who prepared answers to Pearson's end of chapter questions, the Special Master agrees with Chegg that Judge Kiel previously ruled that the names of the individuals who authored the solutions are not relevant. (*See* December 5, 2022 Hearing Transcript at 136:9-137:23; Chegg's April 21, 2023 Supplemental Submission at 5). In addition, during the April 4, 2023 hearing on this topic, Pearson made clear that what it really wants is source information ("I want to know when somebody was sitting down to draft the solution to a question, what they were looking at, how did they draft it"), which relates to and should be covered by the content category described above. (Hrg. at 54:14-55:7). Although the Special Master understands that the Court's prior ruling was without prejudice, Pearson has not articulated a new reason as to why this prior decision of the Court should be revisited at this time.

With respect to Pearson's request for additional fields relating to Chegg's distribution of its answers, Chegg asserts, and the Special Master agrees, that the information Pearson seeks here is best addressed through the user log data Chegg has already been ordered to produce. If, after reviewing the user log data production, Pearson believes it is entitled to additional information on this topic, the parties shall meet and confer before raising any new issue with the Special Master.

With respect to Pearson's request for additional date/timeline related fields, Chegg has agreed to produce data from the additional date fields identified by Pearson in its April 14, 2023

Supplemental Submission (at page 4) to the extent they exist and have been used by Chegg. (See Chegg's April 21, 2023 Supplemental Submission at 6).

Finally, Chegg has agreed to produce data from 19 additional fields to the extent in use, as set forth in its April 21, 2023 Supplemental Submission (Section III, page 6).

### **III. ORDER**

**IT IS** on this 5<sup>th</sup> day of June, 2023,

**ORDERED** that Pearson's application to compel Chegg to produce data from additional fields from Chegg's TBS Database is granted in part and denied in part; and it is further

**ORDERED** that Chegg shall produce data from the nine additional fields from the TBS Database regarding quality control set forth in its April 21, 2023 Supplemental Submission at page 4, to the extent those fields exist and have been used by Chegg; and it is further

**ORDERED** that Chegg shall produce data from five additional fields identified by Pearson in its April 14, 2023 Supplemental Submission at page 3 regarding content used by Chegg in preparing its solutions to Pearson's questions, to the extent those fields exist and have been used at any time by Chegg; and it is further

**ORDERED** that Pearson's request for identifying information of individuals who prepared answers to Pearson's end of chapter questions is denied; and it is further

**ORDERED** that Pearson's request for additional fields relating to Chegg's distribution of its answers is denied; and it is further

**ORDERED** that Chegg shall produce data from the additional date/timeline fields identified by Pearson in its April 14, 2023 Supplemental Submission (at page 4), to the extent they exist and have been used by Chegg; and it is further

**ORDERED** that Chegg shall produce data from 19 additional fields to the extent in use as set forth in its April 21, 2023 Supplemental Submission (Section III, page 6); and it is further

**ORDERED** that Chegg shall produce the data with respect to the text of Pearson's questions in a "readable" format that will be easier for Pearson to understand; and it is further

**ORDERED** that Chegg shall make the foregoing productions within 21 days of the date of this Order.

**SO ORDERED.**

/s/ Jose L. Linares

Hon. Jose L. Linares, U.S.D.J. (Ret.)